

# QA AND OPTIMIZATION OF THE GREEK ENTRIES IN IATE

## 5th MEETING OF GREEK TRANSLATOLOGISTS

Spiros Doikas  
[ttranslatum.gr](http://ttranslatum.gr)



Aristotle  
University of  
Thessaloniki

School of  
French  
Language and  
Literature

Thessaloniki  
21-23 May  
2015

# TBX FILE PROCESSING

# PARSER DEVELOPMENT

A custom-made **script** was developed especially for this task, i.e. the **conversion of the TBX file to a tab delimited format (TSV)**. Each **language** is converted into **two columns**, one for the **full form** and one for the **abbreviations**. Also there is a column for the **term ID** and the **subject field numbers**.

The **pipe (“|”)** delimiter was used for **synonyms**.

# TBX FILE SOURCE CODE

```
<langSet xml:lang="de">↓
  <tig>↓
    <term>Zuständigkeit der Mitgliedstaaten</term>↓
    <termNote type="termType">fullForm</termNote>↓
    <descrip type="reliabilityCode">3</descrip>↓
  </tig>↓
</langSet>↓
<langSet xml:lang="el">↓
  <tig>↓
    <term>αρμοδιότητα των κρατών μελών</term>↓
    <termNote type="termType">fullForm</termNote>↓
    <descrip type="reliabilityCode">3</descrip>↓
  </tig>↓
</langSet>↓
<langSet xml:lang="en">↓
  <tig>↓
    <term>competence of the Member States</term>↓
    <termNote type="termType">fullForm</termNote>↓
    <descrip type="reliabilityCode">3</descrip>↓
  </tig>↓
```

# TBX CONVERTED TO EXCEL SPREADSHEET

Struktural SAP	πρόγραμμα διαρθρωτικής προσαρμογής	ΠΔΠ	structural adjust SAP
prioritäres Veröffen	πρόγραμμα δημοσίευσης με προτεραιότητα	ΠΔΠ	priority publicati PPP
	Πρόγραμμα δημοσίευσης με προτεραιότητα	ΠΔΠ	
Finanzrah MFR	πολυετές δημοσιονομικό πλαίσιο	ΠΔΠ	financial framew FF   MAFF
Aktionsrah AAK	πλαίσιο δράσης για την προσαρμογή	ΠΔΠ	Framework for A FAA
multinationale Schu	Πολυεθνική Δύναμη Προστασίας	ΠΔΠ	Multinational Pr MNPF
Abfälle au HZVA	παραγωγή, διαμόρφωση, προμήθεια κ	ΠΔΠΧ	manufacture, for MFSU
Weltfunkl WRC	Παγκόσμια Διάσκεψη Ραδιοεπικοινωνιών	ΠΔΡ	World Radiocom WRC
Zollzuschl ZZu	πρόσθετος δασμός στη ζάχαρη	πδσ	additional duty cads
Hochschul HKP	διαπανεπιστημιακό πρόγραμμα συνεργασίας	ΠΔΣ	Inter-university ICP
	Περιφερειακό Διοικητικό Συμβούλιο	ΠΔΣ	
Handelspl MTF	μηχανισμός εμπορίας   πολυμερής διευκ	ΠΔΣ   ΠΜΔ	Alternative Trade MTF
Technologiemanage	πρόγραμμα διαχείρισης της τεχνολογίας	ΠΔΤ	technology mana TMP
Technologie TMP	πρόγραμμα διαχείρισης της τεχνολογίας	ΠΔΤ	
Interimsindex der V	προσωρινός δείκτης τιμών καταναλωτή	ΠΔΤΚ	interim index of IICP
Bescheinig CRS	πιστοποιητικό διάθεσης σε υπηρεσία   π	ΠΔΥ	MRC   certificate CRS
	Παράλληλες Επενδύσεις	ΠΕ	
	Πανεπιστημιακή Εκπαίδευση	ΠΕ	
	Περιβαλλοντική Εκπαίδευση	ΠΕ	
	Πανεπιστημιακή Εκπαίδευση	ΠΕ	
Elektrone EI	ιον(τ)ισμός από πρόσπτωση ηλεκτρονίου	ΠΕ	electron-impact EI

# FILE PREPARATION

Once the tab delimited file was created, the following actions were taken:

- Sort on the Greek full-form column
- Removal of in-term quotation marks
- Excel import

# TYPES OF ISSUES

The issues identified can be classified as follows:

- Technical
- Linguistic
- Lexigographic

The emphasis on this paper is on **technical** and **linguistic** issues which are **amenable** to (varied degrees of) **automation**.

# TECHNICAL AND LINGUISTIC ISSUES





1

- Invalid xml characters

As a result of this, the import into xml/tbx processing tools fails

```

9917261 <descrip type="reliabilityCode">3</descrip>↓
9917262 /tig>↓
9917263 angSet>↓
9917264 ngSet xml:lang="el">↓
9917265 tig>↓
9917266 <term>Σχέδιο &#1;δράσης&#1; της &#1;ΕΕ &#1;για&#1; το &#1;ρόλο&#1; της &#1;ισότ<
9917267 <termNote type="termType">fullForm</termNote>↓
9917268 <descrip type="reliabilityCode">3</descrip>↓
9917269 /tig>↓
9917270 tig>↓
9917271 <term>Σχέδιο δράσης για θέματα φύλου</term>↓
9917272 <termNote type="termType">fullForm</termNote>↓
9917273 <descrip type="reliabilityCode">3</descrip>↓
9917274 /tig>↓
9917275 angSet>↓
9917276 ngSet xml:lang="en">↓
9917277 tig>↓
9917278 <term>EU Plan of Action on Gender Equality and Women's Empowerment in Developm<
9917279 <termNote type="termType">fullForm</termNote>↓
9917280 <descrip type="reliabilityCode">4</descrip>↓
9917281 /tig>↓
9917282 tig>↓

```

The text as  
displayed in tbx  
(above) and MS  
Word (right)

Σχέδιο δράσης της ΕΕ για το ρόλο της ισότητας  
των φύλων και της χειραφέτησης των γυναικών  
στην ανάπτυξη





2

- Use of Latin characters instead of Greek ones in Greek words

This causes issues with spell-checking and term recognition in CAT tools' workflow.



Instead of "p" it should be "π".

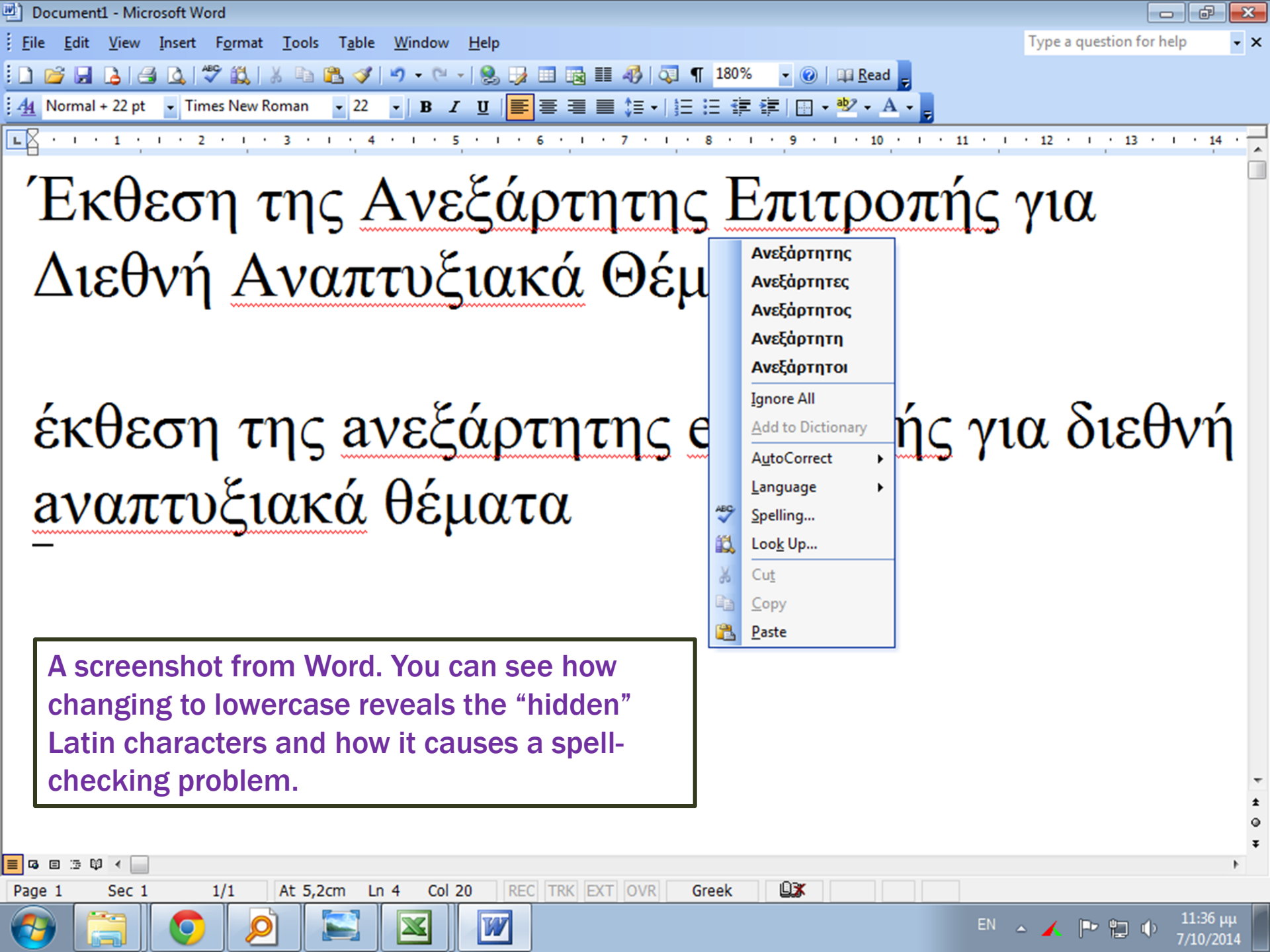
rudder in neutral

en > el (domain: Any domain, type of search: All)

Space transport [CdT]

EN rudder in neutral

EL ρηδάλιο διεύθυνσης σε ουδέτερη θέση



Εκθεση της Ανεξάρτητης Επιτροπής για  
Διεθνή Αναπτυξιακά Θέματα

έκθεση της ανεξάρτητης επιτροπής για διεθνή  
αναπτυξιακά θέματα

A screenshot from Word. You can see how changing to lowercase reveals the “hidden” Latin characters and how it causes a spell-checking problem.



3

- Non-combined diacritics (i.e. accent before the vowel)

This causes issues with spell-checking and term recognition in CAT tools' workflow.



# LATIN CATITAL “E” INSTEAD OF GREEK “Ε”

Original text	Changed to lowercase
Εγχρωμος διαφανής υαλοπίνακας έγχρωμο διαφανές τζάμι	εγχρωμος διαφανής υαλοπίνακας έγχρωμο διαφανές τζάμι
Εγχρωμος υαλοπίνακας	εγχρωμος υαλοπίνακας
Εκθεση της Ανεξάρτητης Επιτροπής για Διεθνή Αναπτυξιακά Θέματα	εκθεση της ανεξάρτητης επιτροπής για διεθνή αναπτυξιακά θέματα
Εκθεση του Εκτελεστικού Συμβουλίου	εκθεση του εκτελεστικού συμβουλίου
Έλεγχος Ειδικής Μεταφοράς	ελεγχος ειδικής μεταφοράς
Έλεγχος ανοπτήσεως	ελεγχος ανοπτήσεως
Έλεγχος δια επαφής	ελεγχος δια επαφής
Έλεγχος δι'αυτοκλείστου	ελεγχος δι'αυτοκλείστου
Έλεγχος κάμψης τριών σημείων	ελεγχος κάμψης τριών σημείων
Έλεγχος με πολαροσκόπιο	ελεγχος με πολαροσκόπιο
Έλεγχος πίεσεως θραύσεως	ελεγχος πίεσεως θραύσεως
Έλεγχος προσβολής άκρων(καθρέπται)	ελεγχος προσβολής άκρων(καθρέπται)
Έλεγχος ρυπάνσεως ατμοσφαιράς	ελεγχος ρυπάνσεως ατμοσφαιράς
Έλεγχος σταθερότητας του φωτός	ελεγχος σταθερότητας του φωτός
Έλεγχος συνολικής μετακίνησης	ελεγχος συνολικής μετακίνησης
Έλεγχος συσκευασμένων	ελεγχος συσκευασμένων
Έλεγχος της γωνίας φάσεως	ελεγχος της γωνίας φάσεως
Έλεγχος ώμου	ελεγχος ώμου

**By changing the text to lowercase, issues where Latin characters are used instead of Greek ones, become visible.**



4

- **Incorrect or no accents**

Words that should have an accent but do not (and vice versa) as well as other types of incorrect accent usage.



## Terms that should *not* have an accent

Χαλασμένα **μπόρ** | αλλοίωση άκρων (edge creep|edge flake)

**τέστ**

(μικροβιολογικό **τέστ** για την ηπατίτιδα)

τείζερ|όπλο εκτόξευσης βελών που προκαλούν ηλεκτρικό **σόκ**

**Μή** κυκλικό άνοιγμα

ντου**ύζ**

τζαμωτό χώρισμα για ντου**ύζ**

χρυσός **μάτ**

αμφίδρομη διαδρομή Γ**ή**-Διάστημα

ανατολή της γ**ής**

Κολέγιο της Ευρώπης στη Μπρ**ύζ**

μπλ**όκ**

τσ**όκ** μαστού

αί**ξ** Angora|κατσίκα Angora

αί**ξ** η ελικόκερως του Σουλεϊμάν

αί**ξ** η ελικόκερως|μαρκόρ

αί**ξ** του Cachemire|κατσίκα του Cachemire

Ευρωπαϊκό Κέντρο για τα Δικαιώματα των Ρ**όμ**

## Terms that *should* have an accent

Αγγιστρο υαλοθραύσματος

Ιδρυμα Ασφαλίσεως

τιθεται σφραγιδα και υπογραφη

Έλεγχος δια επαφής [at least 230 instances of δια]

Ενωση [at least 257 instances of Ενωση – i.e. Διεθνής Ενωση Διερμηνέων  
Συνεδριάσεων]

Υπατος

## Other accentuation issues

υαλούφασμα

υαλο-ύφασμα

ράϊσμα

άϋλες υπηρεσίες

Γκλέϊ

Επιθεώρηση Ναυσιπλοΐας

Δίκτυο

CTS, ελεύθερο για μετάδοση

ακτοπλοΐα [200+ instances]



5

- Spelling errors

Various spelling errors and typos.

## Spelling errors and typos

κοματιάζω γυαλί

άγγιστρο χειριστού

κορδέλλα

τουαλέτιες

ρωσικό

τρελλά

σελλίνια

ανταποδειξη [Also, the Greek “v” is in fact an English lowercase “V”]

διάταξη ενργειακής μετατροπής

σαπωνίνες κουϊλαϊας

αλυσσίδα

διαχείρηση συστήματος πληροφοριών

ανατανακλαστικόν των BECHTEREW-MENDEL

αύξηση της συμμετοχής γυναικών' αύξητη της παρουσίας γυναικών



6

- Wanton use of capitals

Use of capitals for no reason, often with the same term as a synonym in lowercase.

Search


Result 1- 4 of 4 for ατμοσφαιρική ρύπανση

[Full entry](#)

★★★★★ → @



[Full entry](#)



Full  
entry

★★★★☆

★★★★★ → @

★ ★ ★



.....

Full  
entry

★★高中 → @

★★高中 → @

★★高商 → @



7

- **Erroneous delimitation**  
By using arbitrary characters like slash, numbers, alphabetic characters, semicolon, Greek semicolon, apostrophe, dot, “ή”.



# NUMERICAL DELIMITATION

1) δοκιμαστική εφαρμογή 2) (δικαστική) υπόθεση που δημιουργεί προηγούμενο				
1) έκθεση σε χημικά μέσα 2) έκθεση σε χημικούς παράγοντες				
1) εμπόριο εμβρύων 2) ανταλλαγή γονιμοποιημένων ωαρίων				
1) εξαρτημένοι εργαζόμενοι 2) μισθωτοί				
1) εξορθολογισμός της βιομηχανικής παραγωγής 2) ορθολογική διάρθρωση της βιομηχανικής παραγωγής				
1) ισορροπία δικαιωμάτων και υποχρεώσεων 2) εξισορρόπηση ... (ΔΣ/ΑΤ)				
1) μετατοπίσεις της χρήσης της γης 2) μετατοπίσεις των καλλιεργειών				
1) οι οργανισμοί που δηλώνονται (από τα Κράτη μέλη) 2) οι οργανισμοί που ορίζονται 3) οι οργανισμοί που έ				
1) πελάτις επιχείρηση 2) επιχείρηση που κάνει χρήση...				
1) πραγματογνωμοσύνη μισθίου 2) τεχνική περιγραφή της υπάρχουσας κατάστασης				
1) ράγες 2) καρποί χωρίς πυρήνα				
1) ταξίδια κλειστών θυρών 2) περιήγηση κλεισμένων των θυρών				
1) χωρητικότητα των δεξαμενών αποθήκευσης κοπριάς 2) δυναμικότητα αποθήκευσης				
1)εκτελώ μικρές ρυθμίσεις;2)πραγματοποιώ ελάσσονες ρυθμίσεις;εκτελώ μικρορρυθμίσεις				
1)εμφανίζω υποβάθμιση της απόδοσης;2)εμφανίζω υποβάθμιση επιδόσεων				
1)εντείνω τη χρηματοδοτική δραστηριότητα;2)κλιμακώνω τη χρηματοδοτική δραστηριότητα				
1)εξαλείφω υπολείμματα λίπους;2)απομακρύνω κατάλοιπα λίπους;αφαιρώ υπολείμματα λίπους				
1)εξυπηρετώ το χρέος;2)διασφαλίζω την εξυπηρέτηση του χρέους				
1)επαναχορηγώ πόρους;2)κατανέμω κεφάλαια				

# ALPHABETIC DELIMITATION

α)παράγωγο κρυπτογράφημα   β)κρυπτογράφημα αντικατάστασης				
α)περιορισμοί επεξεργασίας   β)επεξεργασία προσωπικών δεδομένων για τρίτους				
α)περιοχές προστασίας   β)πιστοποίηση λογισμικού				
α)περίσφιγξη   β)τοπική όπλιση περιοχής αγκυρώσεως τένοντα   γ)έλασμα ανοξείδωτου χάλυβα				
α)πιστοποίηση λειτουργικού ελέγχου   β)ορθότητα συστήματος				
α)πλήρης απομόνωση   β)περιορισμός προγράμματος				
α)πλήρης μεσολάβηση   β)βλάβη υλικού				
α)προδιαγραφές προστασίας   β)εφαρμογή συστήματος προστασίας				
α)προνομιακή πληροφορία   β)αρχή ελάχιστου προνόμιου				
α)προοδευτική κρυπτογράφηση   β)κρυπτογράφηση καθ'ομάδες				
α)προσπέλαση με βάση την ικανότητα   β)καθορισμός διεύθυνσης με βάση την ικανότητα				
α)προστασία ακεραιότητας   β)διασφάλιση συστήματος				
α)προστασία από κοινολόγηση   β)προστασία από τροποποιήσεις				
α)προστασία με βάση την ικανότητα   β)εξουσιοδότηση				
α)προστασία περιεχομένων   β)προστασία συμφραζομένων				
α)προστασία στο επίπεδο του στοιχείου   β)περιοχή προστασίας				
α)προστατευμένο αντικείμενο   β)διασφάλιση σε στάθμη εγγραφής				
α)προσωπικό διασφάλισης   β)συγχρονισμός διεργασιών				
α)πυρήνας συστήματος   β)αθροιστήρας με βάση 2				
α)συμμετρική προστασία   β)συμμετρική διασφάλιση				

# DELIMITATION WITH APOSTROPHE

## **Delimitation with apostrophe**

οδός των Βαλκανίων' Βαλκανική οδός  
θερμοηλεκτρικός σταθμός' θερμικός σταθμός παραγωγής ηλεκτρικής ενέργειας  
Praesidium της Ευρωπαϊκής Συνέλευσης' Προεδρείο της Ευρωπαϊκής Συνέλευσης  
διέλευση' διάβαση  
αποκατάσταση' επανόρθωση  
μοντέλο πελάτη/εξυπηρετητή' μοντέλο πελάτη/διακομιστή  
κεντρική ιδέα' γενική ιδέα  
προσαράζω' εξοκέλλω  
ΚΕΑ' ΝΑC  
επιτροπή για την κατάρτιση νομοθετικού κειμένου' συντακτική επιτροπή  
βραχύκανο όπλο' πιστόλι' περίστροφο  
τραπεζική εντολή' ένταλμα πληρωμής  
αύξηση της συμμετοχής γυναικών' αύξητη της παρουσίας γυναικών



8

- Broken characters

Characters which broke as a result of unsuccessful conversions between various encodings.

# GREEK CAPITAL “A” WITH TONOS BROKEN

¶λλα απόβλητα		
¶λλα απόβλητα από κατεργασία αερίου		
¶λλα απόβλητα εκρηκτικά		
¶λλα απόβλητα που περιέχουν ανόργανες χημικές ουσίες, π.χ χημικά ερ		
¶λλα απόβλητα που περιέχουν, οργανικές χημικές ουσίες, π.χ χημικά ερ		
¶λλα απόβλητα των οποίων η συλλογή και διάθεση υπόκεινται σε ειδικέ		
¶λλα γαλακτώματα		
¶λλα δημοτικά απόβλητα		
¶λλα έλαια μηχανής κιβωτίου ταχυτήτων και λίπανσης		
¶λλα ιζήματα διύλισης αποστακτήρα και κατάλοιπα αντιδράσεων   ¶λλα		
¶λλα μέταλλα		
¶λλα μη δυνάμενα να λιπασματοποιηθούν απόβλητα		
¶λλα μονωτικά υλικά		
¶λλα πλαστικά		

# “ENTROPY” FULLY BROKEN

← → ↻ 🏠 iate.europa.eu/SearchByQuery.do

cP H 📄 Tr 📄 One 📄 Tr EN-GR 📄 IATE 📄 Gr-En-Mg 📄 En-Gr-Mg 📄 El-Mg 📄 f 📄 in 📄 + Del 📄 o 📄 📄 📄 Anl 📄 [w]

en > el (domain: Any domain, type of search: All)

Result 1- 10 of 10 for Entropy

Electronics and electrical engineering, Physical sciences [COM]		<a href="#">Full entry</a>
EN	<a href="#">entropy</a>	★★★★+@ 📄
EL	εντροπία	★★★★+@
Physical sciences, Chemistry [COM]		<a href="#">Full entry</a>
EN	<a href="#">entropy</a>	★★★★+@ 📄 📖
EL	ἔνζυα	★★★★+@
Information technology and data processing [COM]		<a href="#">Full entry</a>
EN	<a href="#">entropy</a>	★★★★+@ 📄 📖
EL	εντροπία	★★★★+@

# “ENTROPY” IN TBX SOURCE CODE

```
<langSet xml:lang="el">↓
  <tig>↓
    <term>ἔντροπία</term>↓
    <termNote type="termType">fullForm</termNote>↓
    <descrip type="reliabilityCode">3</descrip>↓
  </tig>↓
</langSet>↓
<langSet xml:lang="en">↓
  <tig>↓
    <term>entropy</term>↓
    <termNote type="termType">fullForm</termNote>↓
    <descrip type="reliabilityCode">3</descrip>↓
  </tig>↓
</langSet>↓
```



9

- Cut-off Greek multiword terms

Terms in pieces...



# TERMS IN PIECES

iate.europa.eu/SearchByQuery.do;sessionId=hphrJQhGkyft1F2g41VTKSF7zyvpj9FDM2lYXk6B5rJVε

InterActive Terminology for Europe

πανελλήνια ομοσπονδία εργαζομένων κινηματογράφου, τηλεόρασης και  
el > Any (domain: Any domain, type of search: All)

Search

Result 1- 1 of 1 for πανελλήνια ομοσπονδία εργαζομένων

NO SUBJECT DOMAIN [COM] [Full entry](#)

EL ΠΟΕΚΤ

Πανελλήνια Ομοσπονδία Εργαζομένων Κινηματογράφου, Τηλεόρασης και

★★★★★  
★★★★★ @

Cut-off text, should be:  
Πανελλήνια Ομοσπονδία  
Εργαζομένων  
Κινηματογράφου,  
Τηλεόρασης και  
Συναφών Κλάδων

# TERMS IN PIECES

← → ↺ 🏠 iate.europa.eu/SearchByQuery.do 🔍 ☆ 🛑 🌐

📄 H 📄 Tr 📄 One 📄 Tr EN-GR 📄 IATE 📄 Gr-En-Mg 📄 En-Gr-Mg 📄 EI-Mg 📄 f 📄 in 📄 + Del 📄 📄 📄 📄 📄 Anl 📄 [w] 📄 E-G 📄 WR WR

🌟 🌟 🌟 🌟 🌟 for Europe English (en) Search Screen Help

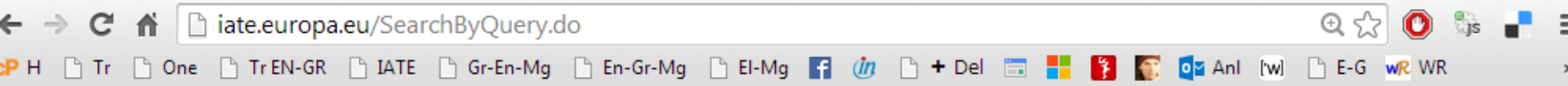
Search

el > Any (domain: Any domain, type of search: All)

Result 1- 3 of 3 for ΕΔΟΕΑΠ

NO SUBJECT DOMAIN [COM]		<a href="#">Full entry</a>
EL	<a href="#">ΕΔΟΕΑΠ</a> Περιθαλψης	★★★★ ★★★★
NO SUBJECT DOMAIN [COM]		<a href="#">Full entry</a>
EL	<a href="#">ΕΔΟΕΑΠ</a> Ενιαίος Δημοσιογραφικός Οργανισμός Επικουρικής Ασφάλισης και	★★★★ ★★★★
Communications systems, FINANCE [EP]		<a href="#">Full entry</a>
EL	<a href="#">ΕΔΟΕΑΠ</a> Ενιαίος Δημοσιογραφικός Οργανισμός Επικουρικής Ασφάλισης και Περιθάλψεως	★★★★ ★★★★ @

# TERMS IN PIECES



English (en)

[Search Screen](#) [Help](#)

ΔΝΣΧΟΠ

Search

el > Any (domain: Any domain, type of search: All)

Result 1- 2 of 2 for ΔΝΣΧΟΠ

NO SUBJECT DOMAIN [COM]

[Full entry](#)

EL ΔΝΣΧΟΠ

Περιβάλλοντος

★★★★

★★★★ @

NO SUBJECT DOMAIN [COM]

[Full entry](#)

EL ΔΝΣΧΟΠ

Διευρυμένο Νομαρχιακό Συμβούλιο Χωροταξίας Οικισμού και

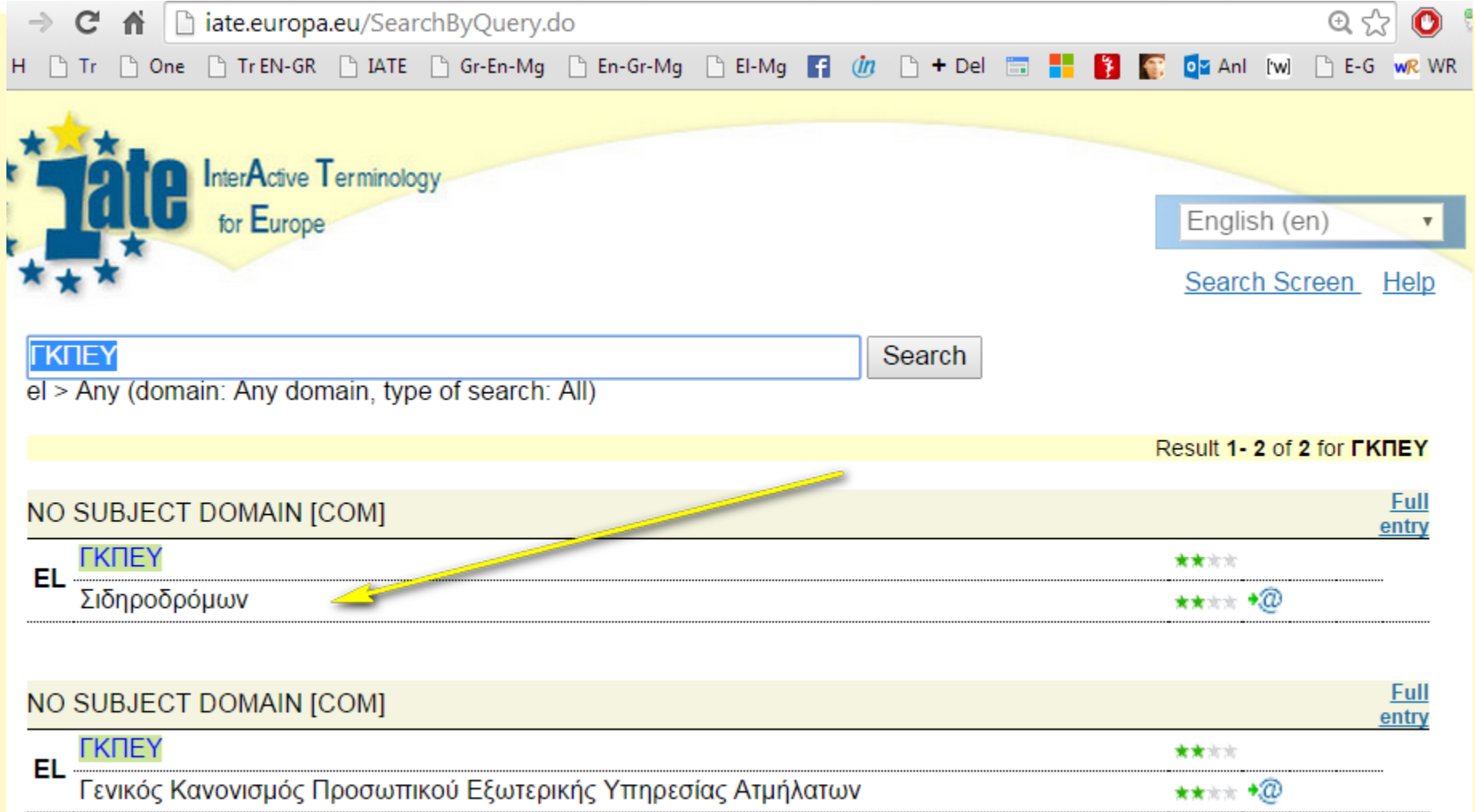
★★★★

★★★★ @

# TERMS IN PIECES

NO SUBJECT DOMAIN [COM]		<a href="#">Full entry</a>
EL	ΕΠΥ Πληροφορικής	★★★★
		★★★★ @
NO SUBJECT DOMAIN [COM]		<a href="#">Full entry</a>
EL	ΕΠΥ Ελληνική Εταιρία Επιστημόνων Ηλεκτρονικών Υπολογιστών	★★★★
		★★★★
Information technology and data processing [EP]		<a href="#">Full entry</a>
EL	ΕΠΥ Ελληνική Εταιρία Επιστημόνων Ηλεκτρονικών Υπολογιστών και Πληροφορικής	★★★★
		★★★★ @
Defence [EP]		<a href="#">Full entry</a>
EL	ΕΠΥ Εθελοντής Πενταετούς Υποχρεώσεως	★★★★
		★★★★ @

## TERMS IN PIECES





10

- Usage of a foreign term in the Greek full form field

Rare, but existent.

# FOREIGN TERM IN THE GREEK FULL FORM FIELD

iate.europa.eu/SearchByQuery.do

InterActive Terminology  
for Europe

English (en)

integração da perspectiva do género  
el > en (domain: Any domain, type of search: All)

Result 1 of 1 for integração da perspectiva do género

Banking [COM]

EL integração da perspectiva do género

EN extract a payment order

Full entry

1

Source language Greek, but actual term is Portuguese



11

- Parenthetical instances of  
...foreign terms or foreign/Greek  
abbreviations in the Greek full form  
field.



) (public analyst)|δημόσιος αναλυτής (  
(chemist (analyst) )|χημικός (αναλυτής)  
Ordinateur au Service de l'Enseignement (OSE)  
κατώτατο όριο (διακύμανσης); limit down  
psilocybe | μανιτάρι

12

- Usage of tags and html entities in-term

## HTML entities

&lt;&lt;ροζ&gt;&gt; υπηρεσίες

&lt;i&gt;cis&lt;/i&gt;-6-δεκαοκτενοϊκό οξύ|πετροσελινικό οξύ

ΑΣ&lt;sub&gt;10&lt;/sub&gt;

αποτύπωμα CO&lt;sub&gt;2&lt;/sub&gt;

## Tags as part of term text

<i/>

<sub>

</sub>

<sup>

</sup>

<D>

<S>

<br>

|  (Ctrl) ▾

13

- Use of subject field, style, and part of speech labels within the full form field

# STYLE LABELS IN FULL FORM FIELD

9	ταμπά	
0	ταχεία απόκριση	
1	τεμπέλης(κοιν.)   αλυσσοτροχός ελευθέρας περιστροφής	
2	τζίνες   κλπ. (δεν υπάρχει συλλογική ονομασία   κόλλες	
3	τηγανίτες πάγου	
4	Τμήμα Χρήστη Τηλεφωνίας	
5	Τμήμα Χρηστών του ISDN	
6	τμηματικές σπουδές	
7	το γόνυ των μεταλλορύχων	
8	το δικαίωμα προτάσεων των νόμων	
9	το παιχνίδι της επιχείρησης	
0	τόμσον	

**Labels describing the term which are attached to the term (no space after the term).**

- Grammatical variants in the full form field, i.e. “-ή, -ό” etc.

## ◀ **Grammatical variants abbreviations**

(-α, -ο)

(-ή, -ό)

(-η, -ο)

, -ή, -ό

-είσα, -έν

### *Examples*

γεννηθείς, -είσα, -έν

Προσανατολισμένος (-η, -ο) σε έργο

15

- Comments, definitions and term source notes in term field.



# COMMENTS, DEFINITIONS AND TERM SOURCE NOTES IN TERM FIELD

## Comments, definitions and term source notes in term field

) (αφορά την αντοχή του υλικού πέραν της ελαστικής περιοχής)|πλαστιμότητας (RF ΟΜΑΔΑ-ΚΑΤΣΑΡΑΚΗΣ - ΕΛΕΤΟ (Ελληνική Εταιρεία Ορολογίας) | μη

## Comments, definitions and term source notes in term field

) (αφορά την αντοχή του υλικού πέραν της ελαστικής περιοχής)|πλαστιμότητας (RF ΟΜΑΔΑ-ΚΑΤΣΑΡΑΚΗΣ - ΕΛΕΤΟ (Ελληνική Εταιρεία Ορολογίας) | μη εφαρμογή της αρχής της εδαφικότητας  
αλόναι (Σπηλιώτης) αλογονωμένος υδρογονάνθρακας (ΑΚ) 3) halon (ΑΚ)  
διαμέρισμα|διοικητική περιφέρεια (σε αντιδιαστολή προς το διαμέρισμα, όταν οι δύο όροι συνυπάρχουν)|τομέας (λ.χ. ταχυδρομικός)

16

- Unnecessary use of ALL CAPS.

# UNNECESSARY USE OF ALL CAPS

- σύνδεσμος του CARCASSONNE ή τρίγωνος σύνδεσμος της ουρήθρας
- ανατανακλαστικόν των BECHTEREW-MENDEL

# **AUTOMATING ERROR FIXING**

# BATCH FIND/REPLACE

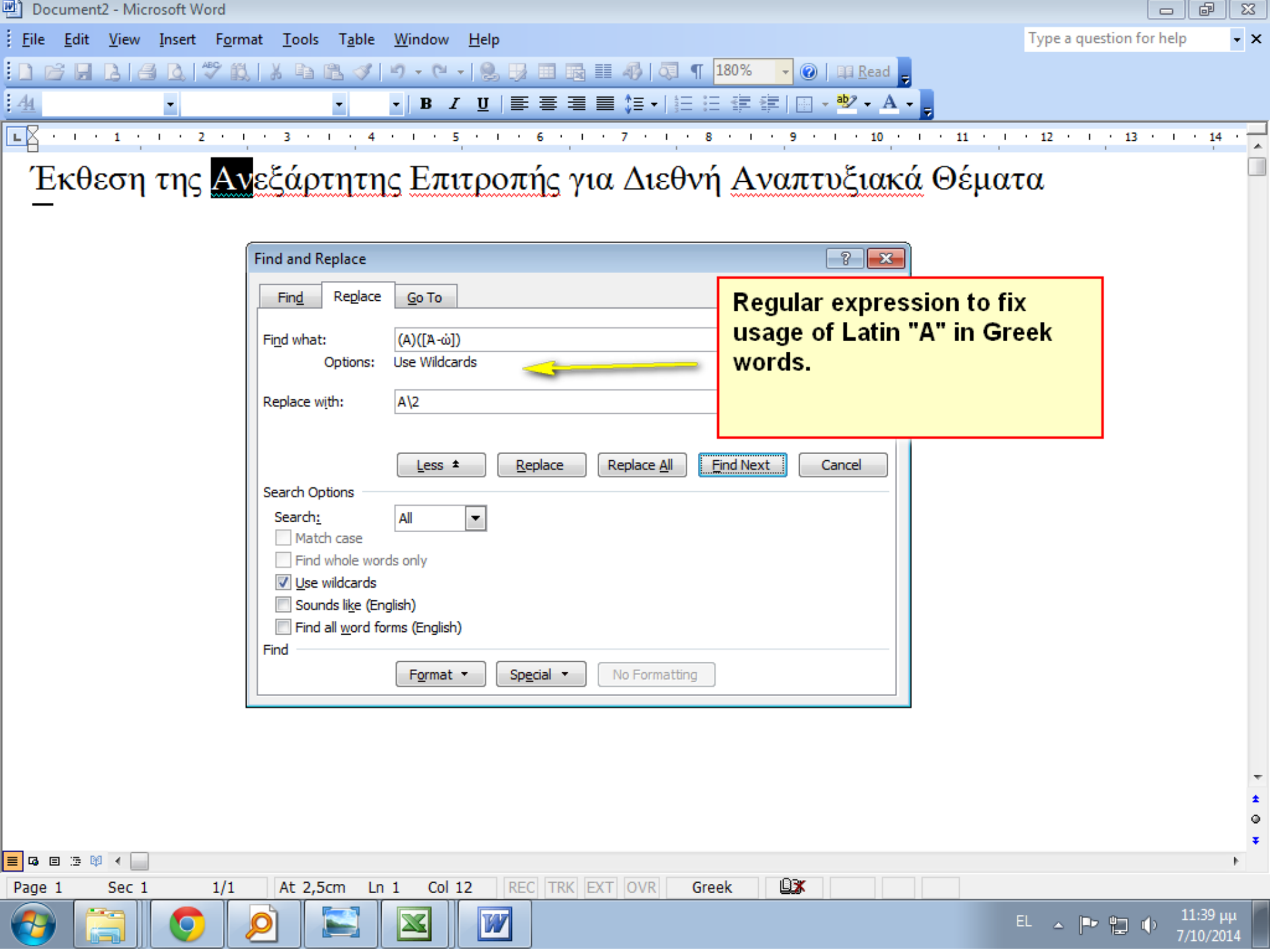
Classification of the most common issues:

- Non-combined diacritics
- Inappropriate use of Latin characters
- Broken characters

See example outcome in [batch-combine-accents.xls](#), an Excel file where a batch/find replace was run.

# REGEX CREATION

Some issues cannot be fixed with a simple **find/replace**, hence appropriate **regexes** need to be created (and possibly be used later as part of a **script** that **automates the corrections** on **database level** or **tbx level**). It would be handy if that script could first **find the matching instances**, and then **perform the replacements** (once approved by terminologists).



Εκθεση της Ανεξάρτητης Επιτροπής για Διεθνή Αναπτυξιακά Θέματα

#### Find and Replace

Find

Replace

Go To

Find what:

(A)([A-ω])

Options:

Use Wildcards

Replace with:

A\2

Less

Replace

Replace All

Find Next

Cancel

#### Search Options

Search:

All

☐ Match case

☐ Find whole words only

☒ Use wildcards

☐ Sounds like (English)

☐ Find all word forms (English)

Find

Format

Special

No Formatting

Regular expression to fix usage of Latin "A" in Greek words.

## **FURTHER WORK AND SUGGESTIONS FOR IMPROVEMENT**

**Apart from some common QA tasks which can be easily automated (like fixing double/multiple/wrong/missing spaces and running a spell check), there are some structural issues which should be taken into account (some of which affect the whole database and not just the Greek part).**



# FURTHER WORK AND SUGGESTIONS FOR IMPROVEMENT

1. Multiple entries for the same source term (i.e. in different subject fields), leading to repetitions and multiple translations (**merge them**)
2. Entries with inconsistent use of *katharevousa* and grammatical rules, i.e. final **-v** (**normalize them**)
3. Descriptive and other comments in parenthesis (**remove/move to appropriate fields**)
4. Usage of various characters to delimit synonyms within the term field (**identify synonyms and move them to a different field**)
5. New terms added should adhere to a strict style guide and terminographic rules to ensure future conformity (**compile style guide and rules?**)

File Home View Add-Ins Help

EN-US

Reverse

EL

Languages

Search Filters Editing Input Model Display Navigation


## Terms

Hitlist


- board
- board
- board
- board
- board cooling
- board crossing
- board drag
- board drop stamp
- board faced grooved timber
- board for pressing
- board for score-opening
- board in the flat
- board mill
- board nail
- board of appeal
- Board of Appeal of the European Chemicals Agency
- Board of Directors of the EBRD
- Board of Directors of the European Investment Bank
- Board of Directors of the World Bank
- Board of Examiners for the European Baccalaureate
- Board of Governors of the EIB (European Investment Bank)

board

Entry number: 34345  
Sub: 3236, 6826

 **EN-US**

**board**

 **EL**

**πλακέτα**

Ver-EN: 2  
Ver-EL: 2

ABC Browse Hitlist Termbases

Terms

Termbase Management

## MULTITERM IMPORT

An example  
of multiple  
instances of  
the same  
term (**board**).

# CONCLUSION

The IATE database is a **precious tool** for translators. By virtue of **EU's extraversion** it is also **downloadable**. However, a lot can be done to **drastically improve and enrich it**, with **limited expenditure**, in comparison to the funds allocated to other translation-related activities and events, and with **exponential benefits**, not only for the EU, but for **translators too**, the world over.

[illegible]